

**Appendix S3.** Assessment of primary study outcomes using adaptation of Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework for assessing the quality of the evidence across studies

**Preterm birth <37 weeks**

Profile of individual studies		Comments
Number of studies	16	<ul style="list-style-type: none"> <li>References: <sup>1–16</sup></li> <li>There are 16 studies but 17 estimates because one study <sup>7</sup> reported separate estimates for 2009 A/pH1N1 and 2008–2009.</li> <li>Additional details on each study can be found in Table S4</li> </ul>
Number of estimates	17	
Estimated number of participants	24,616,017	<ul style="list-style-type: none"> <li>Two studies<sup>5,12</sup> studies counted the total number of birth hospitalizations and the number of antenatal undelivered hospitalizations without reconciling admissions by unique pregnancy, so the counts are overestimated.</li> </ul>
Total number of preterm births	1,207,209	<ul style="list-style-type: none"> <li>This is an underestimate since one study<sup>12</sup> did not provide information on the number of preterm births in their sample.</li> </ul>
<b>Univariable results</b>		
Number of significant effect estimates > 1	2	
Number of non-significant effect estimates	9	
Number of significant effect estimates < 1	0	
Not reported	6	
<b>Multivariable results</b>		
Number of significant effect estimates > 1	4	<ul style="list-style-type: none"> <li>Point estimates ranged from 2.39 to 4.08</li> </ul>
Number of non-significant effect estimates	9	<ul style="list-style-type: none"> <li>Point estimates ranging from 0.82 to 1.27</li> </ul>
Number of significant effect estimates < 1	0	
Not reported	4	
<b>Newcastle Ottawa Score</b>		
0–5	0	
6–7	8	
8–9 (highest quality)	9	<ul style="list-style-type: none"> <li>The RCT<sup>1</sup> is listed with this category but was assessed using the Cochrane Collaboration tool for assessing risk of bias in randomized trials.<sup>17</sup> The risk of bias in the trial was rated as low.</li> </ul>

<b>Risk of diagnostic ascertainment bias (S2 Appendix)</b>		
Very high	2	
High	8	
Medium	6	
Low	1	
<b>Statistical heterogeneity across studies: <math>I^2=97\%</math>; 95% CI: 97–98</b>		
<b>GRADE assessment<sup>a,b</sup></b>		<b>Comments</b>
<b>Phase of investigation</b>	Level 2 (high)	<ul style="list-style-type: none"> <li>We started by assigning a 'high' rating before applying other GRADE criteria since 15 of the 16 studies used cohort designs and sought to confirm the independent association between influenza during pregnancy and preterm birth, and the final study was a randomized clinical trial of influenza vaccination.<sup>18</sup></li> </ul>
<b>GRADE criteria (based on narrative review, not meta-analysis)</b>		
<b>Study limitations:</b> <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> if most evidence is from studies with moderate or unclear risk of bias for most bias domains (serious limitations).</li> <li><b>Downgrade by -2</b> if most evidence is from studies with high risk of bias for almost all bias domains (very serious limitations).</li> </ul>	×	<ul style="list-style-type: none"> <li>Two studies had very high risk of diagnostic ascertainment bias (very serious limitations), and the rest were high or medium (serious limitations).</li> <li><b>Downgrade by 2.</b></li> </ul>
<b>Inconsistency:</b> unexplained heterogeneity or variability in results across studies <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> when estimates of the risk factor association with the outcome vary in direction (for example, some effects appear protective whereas others show risk) and the confidence intervals show no, or minimal overlap.</li> </ul>	×	<ul style="list-style-type: none"> <li>See Forest plot. There is very high heterogeneity in results across studies, qualitatively and quantitatively (<math>I^2=97\%</math>).</li> <li>The confidence intervals of the three studies with the large point estimates do not overlap with most of the null studies.</li> <li>The studies indicating a reduction in preterm birth risk all include the null value, and several are highly imprecise.</li> <li><b>Downgrade by 1.</b></li> </ul>
<b>Indirectness:</b> the study sample, the prognostic factor, and/or the outcome in the primary studies do not accurately reflect the review question <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> when: (1) the final sample only represents a subset of the population of interest; (2) when the complete breadth of the prognostic factor that is being considered in the review question is not well represented in the available studies; or (3) when the outcome that is being considered in the review question is not broadly represented.</li> </ul>		<ul style="list-style-type: none"> <li>No change.</li> </ul>

<b>Imprecision:</b> <ul style="list-style-type: none"> <li>▪ <b>Downgrade by -1</b> if the evidence is generated by a few studies involving a small number of participants and most of the studies provide imprecise results.</li> <li>▪ For narrative summary: <u>Within-study imprecision</u>: (1) sample size justification is not provided and there are less than 10 outcome events for each prognostic variable (for dichotomous outcomes), and (2) no precision in the estimation of the effect size within each primary study. <u>Across study imprecision</u>: there are few studies and small number of participants across studies.</li> </ul>		<ul style="list-style-type: none"> <li>▪ No change.</li> </ul>
<b>Publication bias:</b> <ul style="list-style-type: none"> <li>▪ <b>Downgrade by -1</b> unless the value of the risk/protective factor in predicting the outcome has been repetitively investigated, ideally by phase 2 and 3 studies.</li> </ul>		<ul style="list-style-type: none"> <li>▪ No change.</li> </ul>
<b>Moderate/large effect size:</b> <ul style="list-style-type: none"> <li>▪ <b>Upgrade by +1</b> if moderate or large similar effect is reported by most studies.</li> </ul>		<ul style="list-style-type: none"> <li>▪ No change (most studies do not report large effect sizes and those that do have a high to very high risk of diagnostic ascertainment bias).</li> </ul>
<b>Dose effect:</b> <ul style="list-style-type: none"> <li>▪ <b>Upgrade by +1</b> if possible gradient exists within and between primary studies.</li> </ul>		<ul style="list-style-type: none"> <li>▪ No change.</li> </ul>
<b>GRADE: OVERALL QUALITY OF EVIDENCE</b> <b>(+, very low; ++, low; +++, moderate; +++++, high)</b>	+ <b>Very low</b>	

<sup>a</sup> Based on adaptation<sup>18</sup> of GRADE evaluation framework.<sup>19</sup>

<sup>b</sup> × indicates serious limitations.

### Small-for-gestational-age birth

Profile of individual studies		Comments
Number of studies	5	<ul style="list-style-type: none"> <li>References: 4,16,7,10,20</li> <li>Note there are 5 studies but 6 estimates because one study <sup>7</sup> reported an estimate for 2009 A/pH1N1 and one for the 2008–2009 influenza season.</li> <li>Additional details on each study can be found in Table S7</li> </ul>
Number of estimates	6	
Estimated number of participants	245,127	
Total number of small-for-gestational-age births	20,748	<ul style="list-style-type: none"> <li>Estimated</li> </ul>
<b>Univariable results</b>		
Number of significant effect estimates > 1	1	
Number of non-significant effect estimates	2	
Number of significant effect estimates < 1	0	
Not reported	3	
<b>Multivariable results</b>		
Number of significant effect estimates > 1	2	<ul style="list-style-type: none"> <li>Point estimates ranged from 1.59 to 1.66</li> </ul>
Number of non-significant effect estimates	3	<ul style="list-style-type: none"> <li>Point estimates ranging from 0.71 to 1.14</li> </ul>
Number of significant effect estimates < 1	0	
Not reported	1	
<b>Newcastle Ottawa Score</b>		
0–5	0	
6–7	1	
8–9 (highest quality)	5	
<b>Risk of diagnostic ascertainment bias (S2 Appendix)</b>		
Very high	0	
High	1	
Medium	4	
Low	1	
<b>Statistical heterogeneity across studies: <math>I^2=43\%</math>; 95% CI: 0–76</b>		

GRADE assessment <sup>a,b</sup>		Comments
Phase of investigation	Level 2 (high)	<ul style="list-style-type: none"> <li>These cohort studies were considered to be Level 2 phase studies since they were seeking to confirm the independent association between influenza during pregnancy and small for gestational age.<sup>18</sup></li> <li>Thus, the rating started off as 'high' before application of other GRADE criteria.<sup>18</sup></li> </ul>
<b>GRADE criteria (based on narrative review, not meta-analysis)</b>		
<b>Study limitations:</b> <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> if most evidence is from studies with moderate or unclear risk of bias for most bias domains (serious limitations).</li> <li><b>Downgrade by -2</b> if most evidence is from studies with high risk of bias for almost all bias domains (very serious limitations).</li> </ul>	×	<ul style="list-style-type: none"> <li><b>Downgrade by 1.</b></li> </ul>
<b>Inconsistency:</b> unexplained heterogeneity or variability in results across studies <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> when estimates of the risk factor association with the outcome vary in direction (for example, some effects appear protective whereas others show risk) and the confidence intervals show no, or minimal overlap.</li> </ul>		<ul style="list-style-type: none"> <li>No change.</li> </ul>
<b>Indirectness:</b> the study sample, the prognostic factor, and/or the outcome in the primary studies do not accurately reflect the review question <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> when: (1) the final sample only represents a subset of the population of interest; (2) when the complete breadth of the prognostic factor that is being considered in the review question is not well represented in the available studies; or (3) when the outcome that is being considered in the review question is not broadly represented.</li> </ul>		<ul style="list-style-type: none"> <li>No change.</li> </ul>
<b>Imprecision:</b> <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> if the evidence is generated by a few studies involving a small number of participants and most of the studies provide imprecise results.</li> <li>For narrative summary: <u>Within-study imprecision</u>: (1) sample size justification is not provided and there are less than 10 outcome events for each prognostic variable (for dichotomous outcomes), and (2) no precision in the estimation of the effect size within each primary study. <u>Across study imprecision</u>: there are few studies and small number of participants across studies.</li> </ul>		<ul style="list-style-type: none"> <li>No change.</li> </ul>

<b>Publication bias:</b> <ul style="list-style-type: none"> <li>▪ <b>Downgrade by -1</b> unless the value of the risk/protective factor in predicting the outcome has been repetitively investigated, ideally by phase 2 and 3 studies.</li> </ul>	×	▪ <b>Downgrade by 1.</b>
<b>Moderate/large effect size:</b> <ul style="list-style-type: none"> <li>▪ <b>Upgrade by +1</b> if moderate or large similar effect is reported by most studies.</li> </ul>		▪ No change.
<b>Dose effect:</b> <ul style="list-style-type: none"> <li>▪ <b>Upgrade by +1</b> if possible gradient exists within and between primary studies.</li> </ul>		▪ No change.
<b>GRADE: OVERALL QUALITY OF EVIDENCE</b> <b>(+, very low; ++, low; +++, moderate; +++++, high)</b>	<b>++</b> <b>Low</b>	

<sup>a</sup> Based on adaptation<sup>18</sup> of GRADE evaluation framework.<sup>19</sup>

<sup>b</sup> × indicates serious limitations.

## Fetal death

Profile of individual studies		Comments
Number of studies	10	<ul style="list-style-type: none"> <li>References: <sup>1,5,6,9,14,16,20-23</sup></li> <li>Note that one study <sup>20</sup> had no fetal death events among unexposed women (and only one event among exposed women), and two studies <sup>1,16</sup> reported two estimates, one for abortion, and the second for intrauterine fetal death (definitions were not provided).</li> <li>Additional details on each study can be found in Table S8</li> </ul>
Number of estimates	11	
Estimated number of participants	17,666,768	<ul style="list-style-type: none"> <li>One study<sup>5</sup> study counted the total number of birth hospitalizations and the number of antenatal undelivered hospitalizations without reconciling admissions by unique pregnancy, so the counts are overestimated.</li> </ul>
Total number of fetal deaths	103,902	<ul style="list-style-type: none"> <li>7 out of the 10 studies had fewer than 20 fetal death outcomes in total (8,13,15,18–21)</li> </ul>
<b>Univariable results</b>		
Number of significant effect estimates > 1	3	
Number of non-significant effect estimates	4	
Number of significant effect estimates < 1	0	
Not reported	4	
<b>Multivariable results</b>		
Number of significant effect estimates > 1	3	
Number of non-significant effect estimates	3	
Number of significant effect estimates < 1	0	
Not reported	5	
<b>Newcastle Ottawa Score</b>		
0–5	0	
6–7	6	
8–9 (highest quality)	5	<ul style="list-style-type: none"> <li>The RCT<sup>1</sup> is listed with this category but was assessed using the Cochrane Collaboration tool for assessing risk of bias in randomized trials.<sup>17</sup> The risk of bias in the trial was rated as low.</li> </ul>

<b>Risk of diagnostic ascertainment bias (S2 Appendix)</b>		
Very high	1	
High	3	
Medium	2	
Low	4	
<b>Statistical heterogeneity across studies: <math>I^2=41\%</math>; 95% CI: 0–69</b>		
<b>GRADE assessment<sup>a,b</sup></b>		<b>Comments</b>
<b>Phase of investigation</b>	Level 2 (high)	<ul style="list-style-type: none"> <li>We started by assigning a ‘high’ rating before applying other GRADE criteria since 9 of the 10 studies used cohort designs and sought to confirm the independent association between influenza during pregnancy and fetal death, and the final study was a randomized clinical trial of influenza vaccination.<sup>18</sup></li> </ul>
<b>GRADE criteria (based on narrative review, not meta-analysis)</b>		
<b>Study limitations:</b> <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> if most evidence is from studies with moderate or unclear risk of bias for most bias domains (serious limitations).</li> <li><b>Downgrade by -2</b> if most evidence is from studies with high risk of bias for almost all bias domains (very serious limitations).</li> </ul>	×	<ul style="list-style-type: none"> <li><b>Downgrade by 2.</b></li> <li>Most evidence comes from one study<sup>5</sup> which is very large, but has a very high risk of diagnostic ascertainment bias.</li> </ul>
<b>Inconsistency:</b> unexplained heterogeneity or variability in results across studies <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> when estimates of the risk factor association with the outcome vary in direction (for example, some effects appear protective whereas others show risk) and the confidence intervals show no, or minimal overlap.</li> </ul>		<ul style="list-style-type: none"> <li>No change.</li> </ul>
<b>Indirectness:</b> the study sample, the prognostic factor, and/or the outcome in the primary studies do not accurately reflect the review question <ul style="list-style-type: none"> <li><b>Downgrade by -1</b> when: (1) the final sample only represents a subset of the population of interest; (2) when the complete breadth of the prognostic factor that is being considered in the review question is not well represented in the available studies; or (3) when the outcome that is being considered in the review question is not broadly represented.</li> </ul>	×	<ul style="list-style-type: none"> <li><b>Downgrade by 1.</b></li> <li>Studies used different definitions of fetal death, with differing gestational thresholds and many did not provide a definition.</li> </ul>



<b>Imprecision:</b> <ul style="list-style-type: none"> <li>▪ <b>Downgrade by -1</b> if the evidence is generated by a few studies involving a small number of participants and most of the studies provide imprecise results.</li> <li>▪ For narrative summary: <u>Within-study imprecision</u>: (1) sample size justification is not provided and there are less than 10 outcome events for each prognostic variable (for dichotomous outcomes), and (2) no precision in the estimation of the effect size within each primary study. <u>Across study imprecision</u>: there are few studies and small number of participants across studies.</li> </ul>	×	▪ <b>Downgrade by 1.</b>
<b>Publication bias:</b> <ul style="list-style-type: none"> <li>▪ <b>Downgrade by -1</b> unless the value of the risk/protective factor in predicting the outcome has been repetitively investigated, ideally by phase 2 and 3 studies.</li> </ul>	×	▪ <b>Downgrade by 1.</b>
<b>Moderate/large effect size:</b> <ul style="list-style-type: none"> <li>▪ <b>Upgrade by +1</b> if moderate or large similar effect is reported by most studies.</li> </ul>		▪ No change.
<b>Dose effect:</b> <ul style="list-style-type: none"> <li>▪ <b>Upgrade by +1</b> if possible gradient exists within and between primary studies.</li> </ul>		▪ No change.
<b>GRADE: OVERALL QUALITY OF EVIDENCE</b> <b>(+, very low; ++, low; +++, moderate; +++++, high)</b>	+ <b>Very low</b>	

<sup>a</sup> Based on adaptation<sup>18</sup> of GRADE evaluation framework.<sup>19</sup>

<sup>b</sup> × indicates serious limitations.

## References

1. Madhi SA, Cutland CL, Kuwanda L, Weinberg A, Hugo A, Jones S, et al. Influenza vaccination of pregnant women and protection of their infants. *N Engl J Med* 2014;371:918–31.
2. Ahrens KA, Louik C, Kerr S, Mitchell AA, Werler MM. Seasonal influenza vaccination during pregnancy and the risks of preterm delivery and small for gestational age birth. *Paediatr Perinat Epidemiol* 2014;28:498–509.
3. Doyle TJ, Goodin K, Hamilton JJ. Maternal and neonatal outcomes among pregnant women with 2009 pandemic influenza A(H1N1) illness in Florida, 2009-2010: a population-based cohort study. *PLoS One* 2013;8:e79040.
4. Naresh A, Fisher BM, Hoppe KK, Katov J, Xu J, Hart J, et al. A multicenter cohort study of pregnancy outcomes among women with laboratory-confirmed H1N1 influenza. *J Perinatol* 2013;33:939–43.
5. Martin A, Cox S, Jamieson DJ, Whiteman MK, Kulkarni A, Tepper NK. Respiratory illness hospitalizations among pregnant women during influenza season, 1998-2008. *Matern Child Health J* 2013;17:1325–31.
6. Håberg SE, Trogstad L, Gunnes N, Wilcox AJ, Gjessing HK, Samuelsen SO, et al. Risk of fetal death after pandemic influenza virus infection or vaccination. *N Engl J Med* 2013;368:333–40.
7. Hansen C, Desai S, Bredfeldt C, Cheetham C, Gallagher M, Li DK, et al. A large, population-based study of 2009 pandemic Influenza A virus subtype H1N1 infection diagnosis during pregnancy and outcomes for mothers and neonates. *J Infect Dis* 2012;206:1260–8.
8. Morken NH, Gunnes N, Magnus P, Jacobsson B. Risk of spontaneous preterm delivery in a low-risk population: the impact of maternal febrile episodes, urinary tract infection, pneumonia and ear-nose-throat infections. *Eur J Obstet Gynecol Reprod Biol* 2011;159:310–4.
9. Pierce M, Kurinczuk JJ, Spark P, Brocklehurst P, Knight M. Perinatal outcomes after maternal 2009/H1N1 infection: national cohort study. *BMJ* 2011;342:d3214.
10. McNeil SA, Dodds LA, Fell DB, Allen VM, Halperin BA, Steinhoff MC, et al. Effect of respiratory hospitalization during pregnancy on infant outcomes. *Am J Obstet Gynecol* 2011;204:S54–7.
11. Rogers VL, Sheffield JS, Roberts SW, McIntire DD, Luby JP, Trevino S, et al. Presentation of seasonal influenza A in pregnancy: 2003-2004 influenza season. *Obstet Gynecol* 2010;115:924–9.
12. Cox S, Posner SF, McPheeters M, Jamieson DJ, Kourtis AP, Meikle S. Hospitalizations with respiratory illness among pregnant women during influenza season. *Obstet Gynecol* 2006;107:1315–22.
13. Acs N, Bánhidý F, Puhó E, Czeizel AE. Pregnancy complications and delivery outcomes of pregnant women with influenza. *J Matern Fetal Neonatal Med* 2006;19:135–40.

14. Hartert TV, Neuzil KM, Shintani AK, Mitchell EF, Snowdon MS, Wood LB, et al. Maternal morbidity and perinatal outcomes among pregnant women with respiratory hospitalizations during influenza season. *Am J Obstet Gynecol* 2003;189:1705–12.
15. Tuyishime JD, De Wals P, Moutquin JM, Frost E. Influenza-like illness during pregnancy: results from a study in the eastern townships, Province of Quebec. *J Obstet Gynaecol Can* 2003;25:1020–5.
16. Nieto-Pascual L, Arjona-Berral JE, Marín-Martín EM, Muñoz-Gomariz E, Ilich I, Castelo-Branco C. Early prophylactic treatment in pregnant women during the 2009-2010 H1N1 pandemic: obstetric and neonatal outcomes. *J Obstet Gynaecol* 2013;33:128–4.
17. Higgins JPT, Altman DG, Gøtzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
18. Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. *Syst Rev* 2013;2:71.
19. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction- GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.
20. Irving WL, James DK, Stephenson T, Laing P, Jameson C, Oxford JS, et al. Influenza virus infection in the second and third trimesters of pregnancy: a clinical and seroepidemiological study. *BJOG* 2000;107:1282–9.
21. Stanwell-Smith R, Parker AM, Chakraverty P, Soltanpoor N, Simpson CN. Possible association of influenza A with fetal loss: investigation of a cluster of spontaneous abortions and stillbirths. *Commun Dis Rep CDR Rev* 1994;4:R28–32.
22. Korones SB, Todaro J, Roane JA, Sever JL. Maternal virus infection after the first trimester of pregnancy and status of offspring to 4 years of age in a predominantly Negro population. *J Pediatr* 1970;77:245–51.
23. Wilson MG, Stein AM. Teratogenic effects of asian influenza. An extended study. *JAMA* 1969;210:336–7.